SNS COLLEGE OF TECHNOLOGY

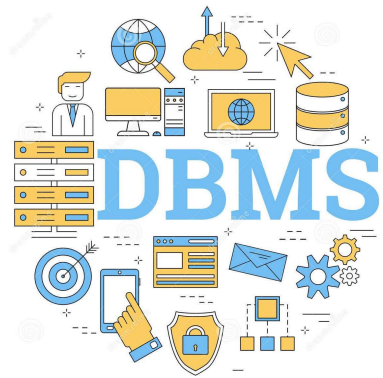**Coimbatore-35**
An Autonomous Institution

Department of Information Technology

19CST202 – Database Management System
II B.Tech. AIML/ IV SEMESTER

UNIT V : PHYSICAL STORAGE AND MONGODB

Topic 2 : File Organization

Data Storage and Indexes – RAID- **File Organization**-Indexing and Hashing –Ordered Indices – B+ tree Index Files – B tree Index Files – Static Hashing – Dynamic Hashing. Query Processing Overview. MongoDB- Installing and Set up, Database creation and manipulation, Indexing and ordering. CASE Studies- Oracle, DB2: Storage and Indexing

Introduction

- A database consists of a huge amount of data. The data is grouped within a table

  in RDBMS, and each table has related records.

What is a File?

A file is named a collection of related information that is recorded on secondary storage such as **magnetic disks**, **magnetic tapes,** and **optical disks**.
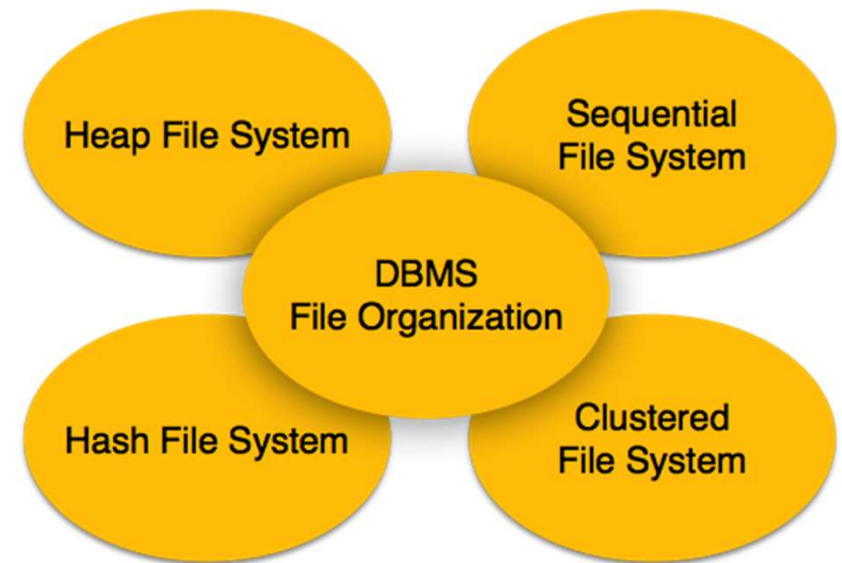
File Organization

- **logical relationships** among various records that **constitute the file**, particularly with respect to the means of <mark>identification and access</mark> to any **specific record.**

- In simple terms, <mark>Storing the files in a certain order</mark> is called File Organization.

- File Structure refers to the <mark>format of the label and data blocks</mark> and of any logical control record.
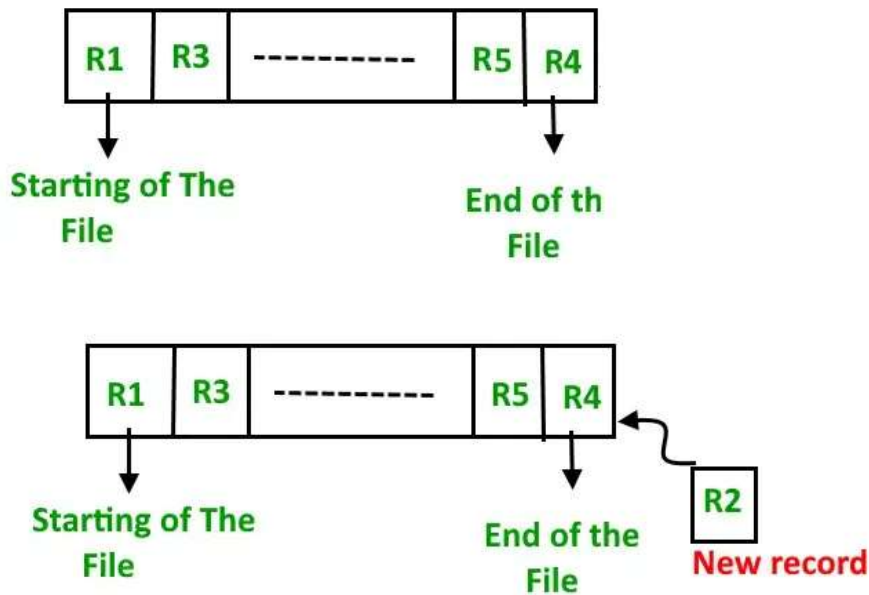
Types of File Organizations

- Sequential File Organization

- Heap File Organization

- Hash File Organization

- B+ Tree File Organization

- Clustered File Organization
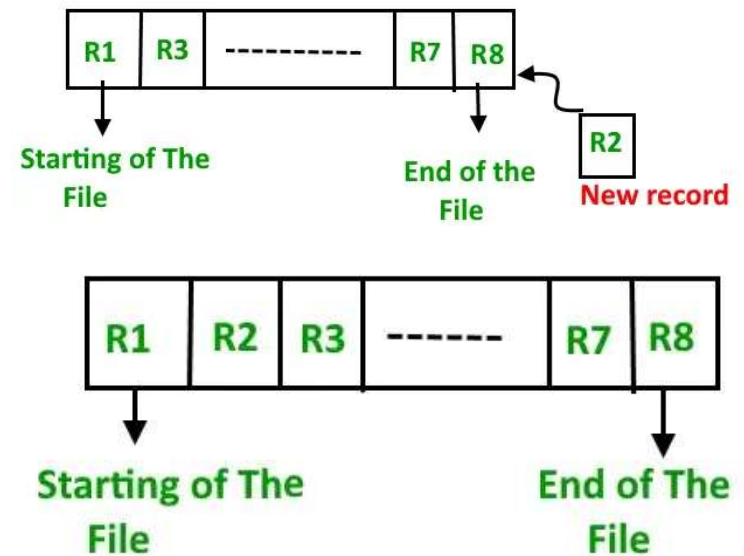
- ISAM (Indexed Sequential Access Method)

Sequential File Organization

The file is stored one after another in a sequential manner.
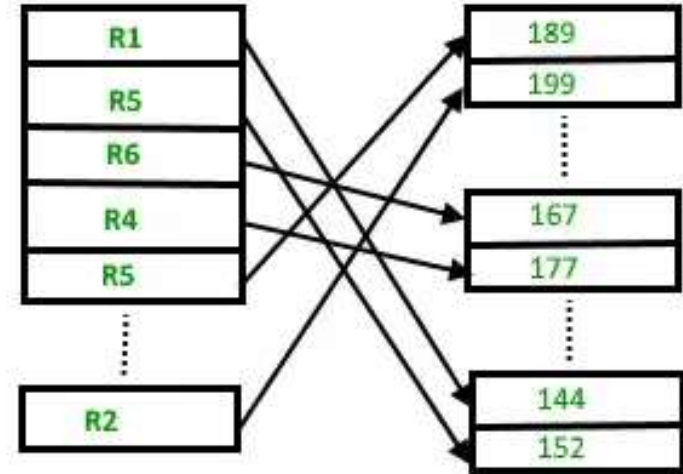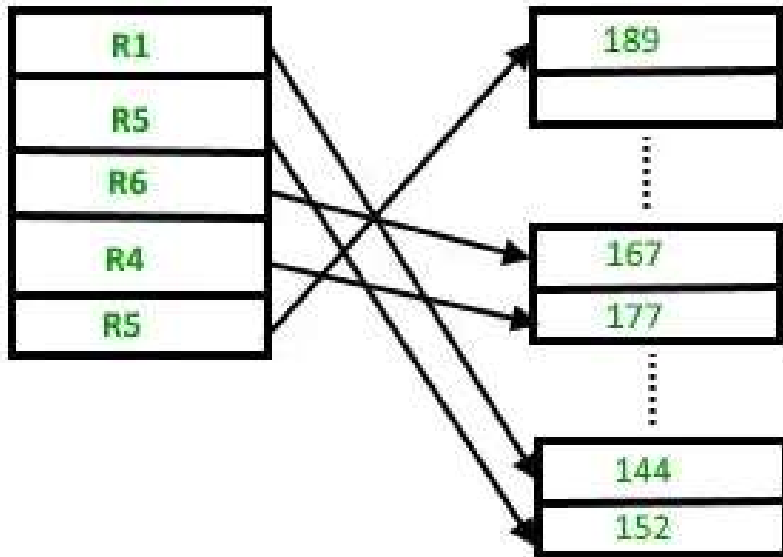
Pile File Method

Sorted File Method

Heap File Organization

- works with data blocks.

- records are inserted at the end of the file, into the data blocks.

- No Sorting or Ordering is required in this method.

- If a data block is full, the new record is stored in some other block,

Heap File Organization

Hash File Organization

- **Hashing** is an efficient technique to <mark>directly search the location of desired data</mark> on the disk **without using an index structure**.

- Data is stored at the **data blocks whose address is generated by using a hash function.**

- The memory location where these records are stored is called a **data block or data bucket.**

Hash File Organization

- **Data bucket –** Data buckets are the memory locations where the records are stored. These buckets are also considered Units of Storage.

- **Hash Function –** The hash function is a mapping function that maps all the sets of search keys to the actual record address. Generally, the hash function uses the primary key to generate the hash index – the address of the data block. The hash function can be a simple mathematical function to any complex mathematical function.

- **Hash Index-**The prefix of an entire hash value is taken as a hash index. Every hash index has a depth value to signify how many bits are used for computing a hash function. These bits can address 2n buckets. When all these bits are consumed? then the depth value is increased linearly and twice the buckets are allocated.
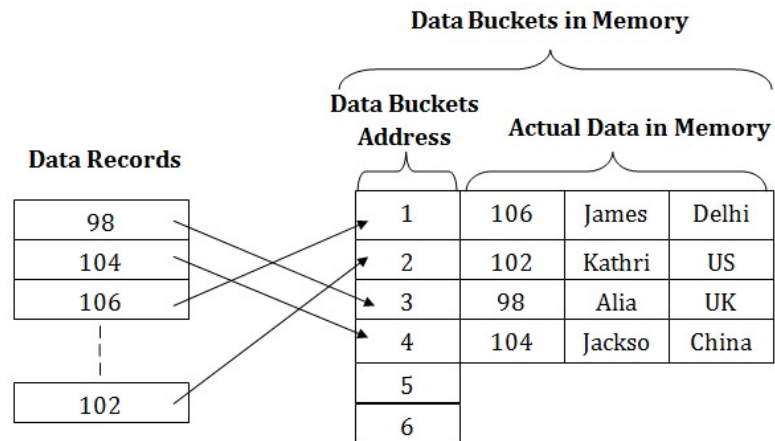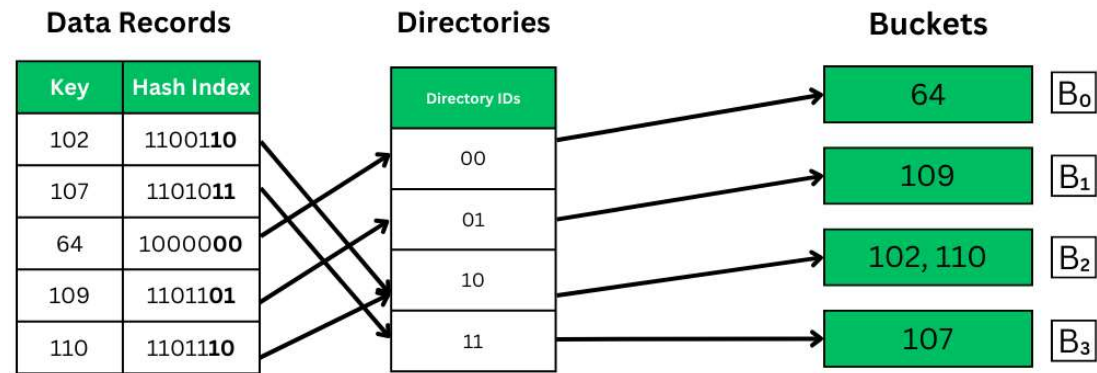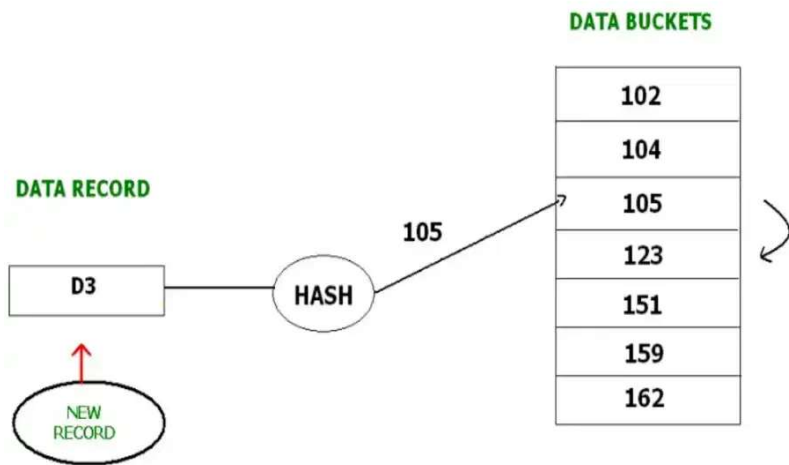

Operations
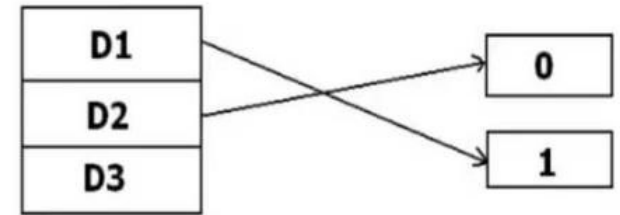- Insertion
- Searching
- Deletion
- Updation

A.Aruna / AP / IT / SEM 4 / DBMS

24-06-2024

TYPES

Static Hashing

Dynamic Hashing

h(D1) -> 1001
h(D2) -> 0101
h(D3) -> 1010

h(D1) -> 1001
h(D2) -> 0101
h(D3) -> 1010

DATA BUCKETS

| 102 |
| 104 |
| 105 |
| 123 |
| 151 |
| 159 |
| 162 |

DATA RECORD

105

D3

HASH

NEW RECORD

D1		0
D2		1
D3		

D1		00
D2		01
D3		10
		11
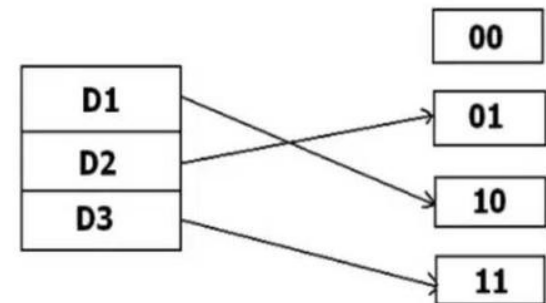
STATIC HASHING	DYNAMIC HASHING
A hashing technique that allows users to perform lookups on a finalized dictionary set (all objects in the dictionary are final and not changing)	A hashing technique in which the data buckets are added and removed dynamically and on demand
Resultant data bucket address is always the same	Data buckets change depending on the records
Less efficient	More efficient

Static Hashing vs Dynamic Hashing

Tree File Organization

B-Tree
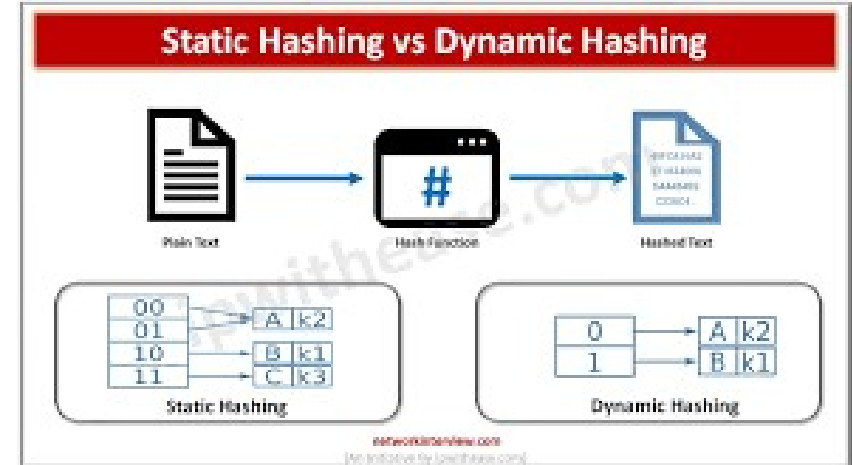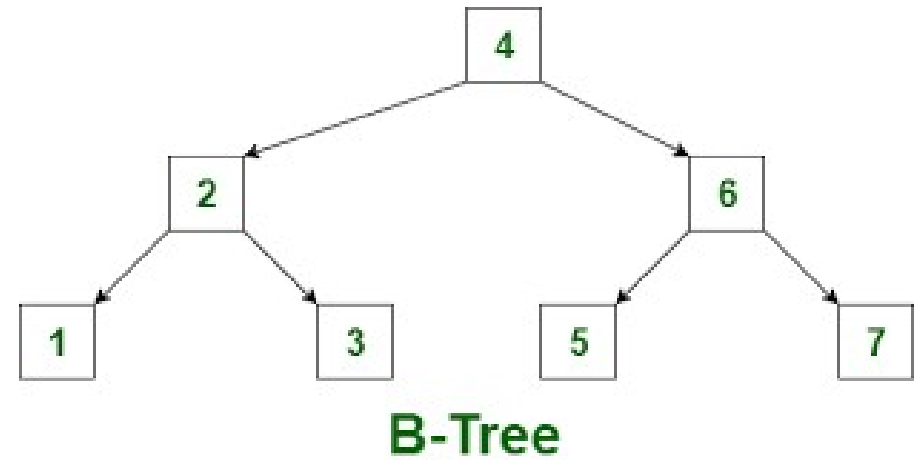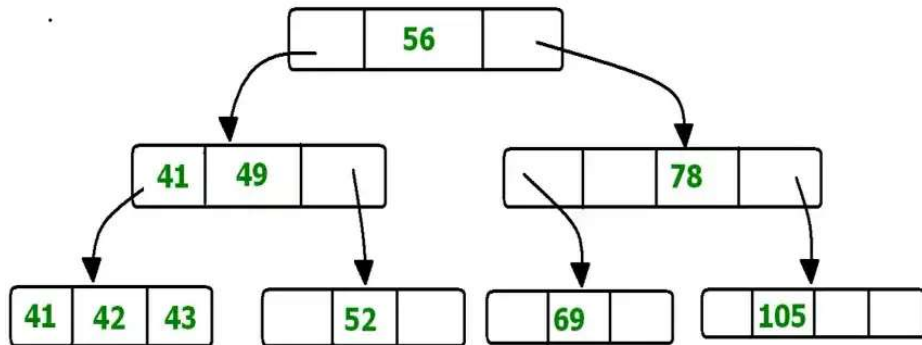
Cluster File Organization

Indexed Clusters and Hash Clusters

EMPLOYEE

EMP ID	EMP_NAME	EMP_ADD	DEP_ID
01	JOE	CAPE TOWN	D_101
02	ANNIE	FRANSISCO	D_103
03	PETER	CROY CITY	D_101
04	JOHN	FRANSISCO	D_102
05	LUNA	TOKYO	D_106
06	SONI	W.LAND	D_105
07	SAKACHI	TOKYO	D_104
08	MARY	NOVI	D_101

DEPARTMENT

DEP_ID	DEP_NAME
D_101	ECO
D_102	CS
D_103	JAVA
D_104	MATHS
D_105	BIO
D_106	CIVIL

CLUSTER KEY

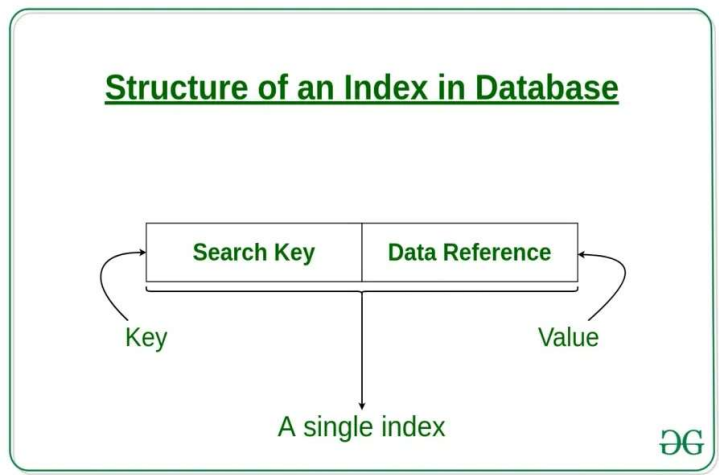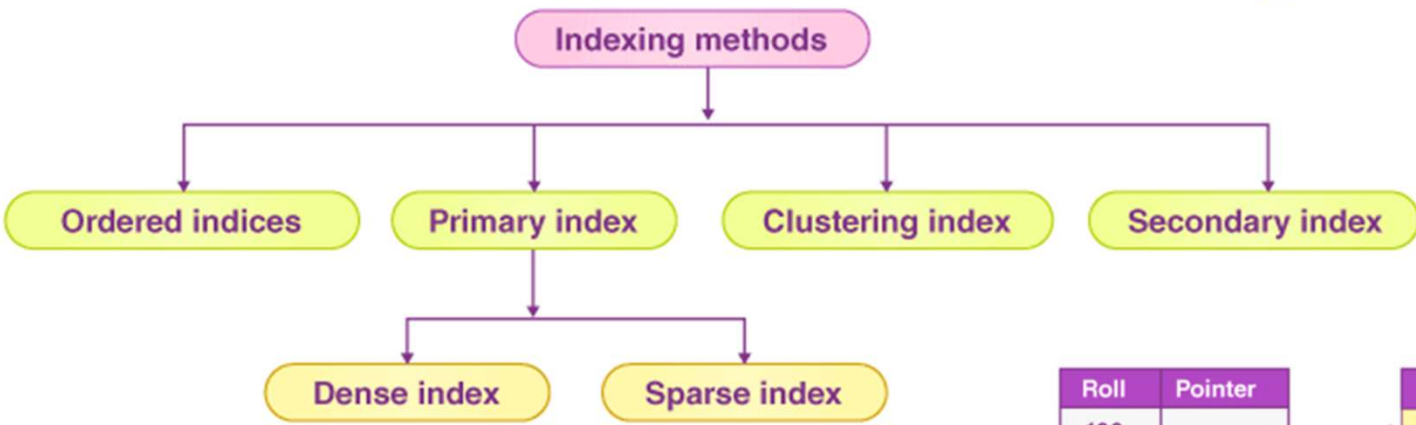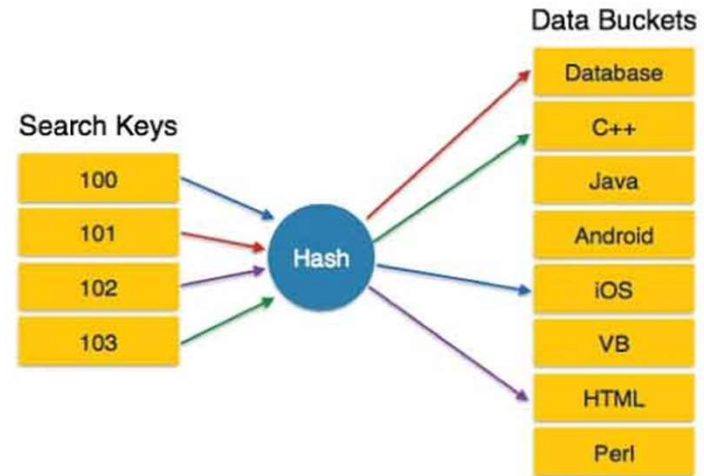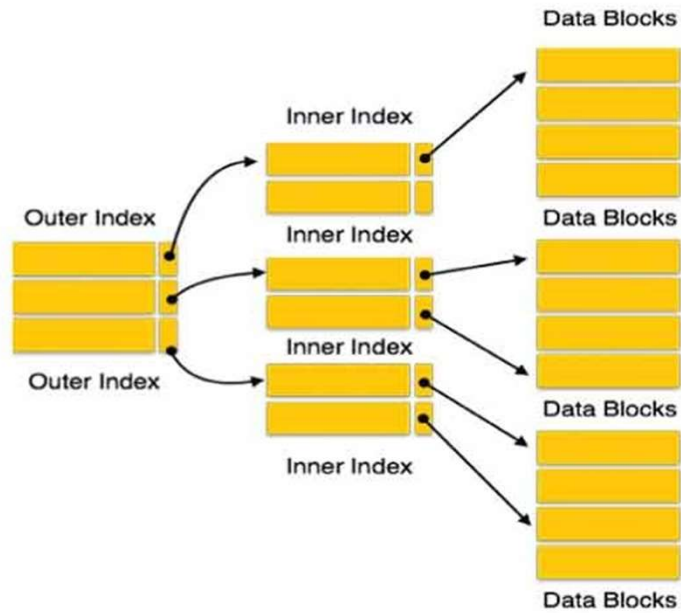| DEP_ID	DEP_NAME	EMP ID	EMP_NAME	EMP_ADD
D_101	ECO	01	JOE	CAPE TOWN
		02	PETER	CROY CITY
		03	MARY	NOVI
D_102	CS	04	JOHN	FRANSISCO
D_103	JAVA	05	ANNIE	FRANSISCO
D_104	MATHS	06	SAKACHI	TOKYO
D_105	BIO	07	SONI	W.LAND
D_106	CIVIL	08	LUNA	TOKYO

DEPARTMENT + EMPLOYEE

ISAM -Indexed Sequential Access Method

- A combination of sequential and indexed methods. Data is stored sequentially, but an index is maintained for faster access.

- Think of it like having **a bookmark in a book** that guides you to specific pages.
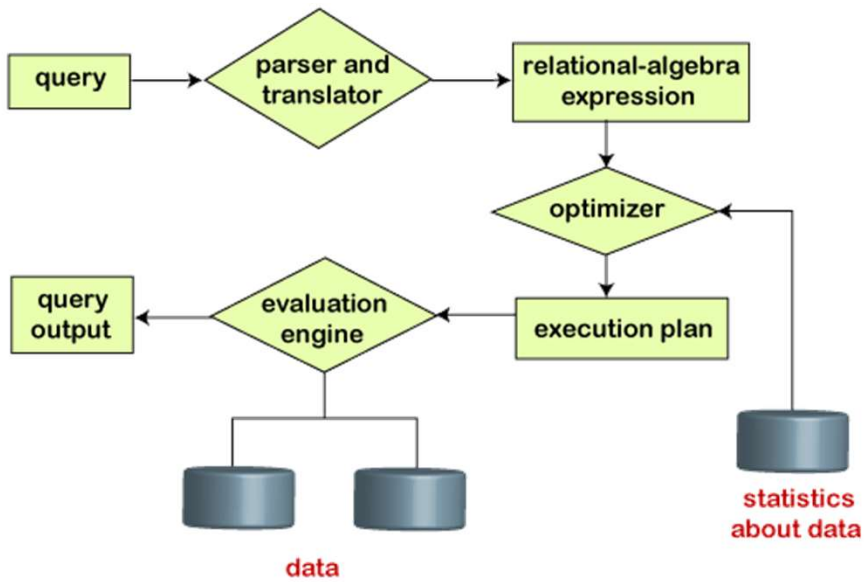
Indexing

Indexing methods
- Ordered indices
- Primary index
  - Dense index
  - Sparse index
- Clustering index
- Secondary index

Structure of an Index in Database

Search Key	Data Reference

Key — Value

A single index

Roll	Pointer
100	
200	
300	

Primary Level Index (RAM)

Roll	Pointer
100	
110	
120	
200	
210	
220	
300	
320	
310	

Secondary Level Index (Hard Disk)

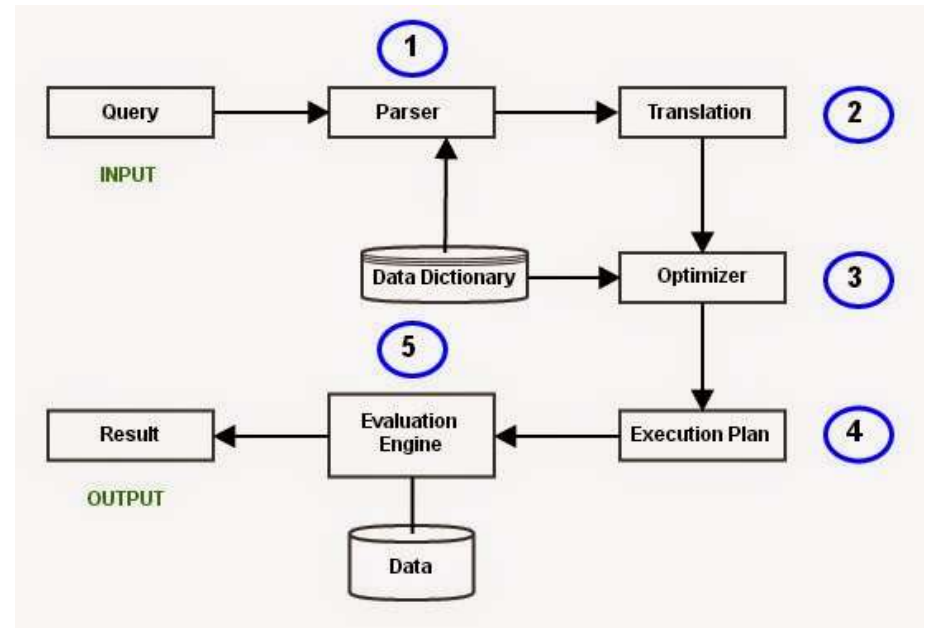| Data bock in Memory	
100	
101	
- - -	- - -
110	
111	
110	- - -
120	
121	
- - -	
200	
201	
- - -	- - -
210	
211	
- - -	- - -
300	
- - -	- - -

Hashing and Indexing

Query Processing Overview

Steps in query processing

TEXT BOOKS

Abraham Silberschatz, Henry F. Korth, S. Sudharshan, —Database System Concepts ٯ , Sixth Edition, Tata McGraw Hill, 2011.

RamezElmasri, Shamkant B. Navathe, —Fundamentals of Database Systems ٯ ,Sixth Edition, Pearson Education, 2011.

Tiwari, Shashank. Professional NoSQL. John Wiley& Sons, 2011

REFERENCES

C.J.Date, A.Kannan, S.Swamynathan, —An Introduction to Database Systems, Eighth Edition, Pearson Education, 2006.
Raghu Ramakrishnan, —Database Management Systems ٯ ,Fourth Edition, McGraw-Hill College Publications, 2015
G.K.Gupta,"Database Management Systems, Tata McGraw Hill, 2011.
S.K.Singh, "Database Systems Concepts, Design and Applications", First Edition,Pearson Education, 2009.